

REPORT DOCUMENTATION PAGE				Form Approved OMB No. 0704-0188	
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.					
1. REPORT DATE (DD-MM-YYYY) 29-Sep-2006		2. REPORT TYPE REPRINT		3. DATES COVERED (From - To)	
4. TITLE AND SUBTITLE IMPROVED EVENT LOCATION UNCERTAINTY ESTIMATES				5a. CONTRACT NUMBER FA8718-05-C-0018	
				5b. GRANT NUMBER	
				5c. PROGRAM ELEMENT NUMBER 62601F	
6. AUTHOR(S) István Bondár, Keith McLaughlin, and Hans Israelsson				5d. PROJECT NUMBER 1010	
				5e. TASK NUMBER SM	
				5f. WORK UNIT NUMBER A1	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Science Applications International Corporation 10260 Campus Point Drive San Diego, CA 92121-1152				8. PERFORMING ORGANIZATION REPORT NUMBER	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Research Laboratory 29 Randolph Road Hanscom AFB, MA 01731-3010				10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/VSBYE	
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-VS-HA-TR-2006-1138	
12. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited.					
13. SUPPLEMENTARY NOTES Reprinted from: Proceedings of the 28 <sup>th</sup> Seismic Research Review – Ground-Based Nuclear Explosion Monitoring Technologies, 19 – 21 September 2006, Orlando, FL, Volume I pp 368 - 376.					
14. ABSTRACT <p>While many recent studies aimed to reduce location bias by introducing improved travel-time corrections, less effort was devoted to the complete estimation of location uncertainty, despite the fact that formal error ellipses are often overly optimistic. Since most location algorithms assume that the observations are independent, correlated systematic errors that are due to similar ray paths inevitably result in underestimated location uncertainties. Furthermore, the tails of real seismic data distributions are heavier than Gaussian. The main objectives of this project are to develop, test, and validate methodologies to estimate location uncertainties in the presence of correlated, systematic, and non-Gaussian errors. Particular attention is paid to robust and transportable models for a travel-time covariance matrix.</p> <p>To address correlated errors, we estimate the spatial correlation structure in arrival-time data using variogram models. We developed a methodology based on copula theory to derive robust, data-driven variogram models. For validation purposes, we use GT0-2 event clusters. These include the Nevada, Lop Nor, Semipalatinsk, and Novaya Zemlya test sites, as well as the Azgir Peaceful Nuclear Explosions and the Lubin, Poland, mine-related events.</p> <p>Using ground-truth (GT) clusters allows us to calculate "ground truth" residuals with respect to the GT locations for a specific velocity model. We show the improvements in variogram estimates when using global 3D models instead of the iasp91 model. The proper choice of the underlying velocity model is especially important for regional phases.</p> <p>To address issues embodied by real data distributions, we performed fully controlled experiments using known high signal-to-noise ratio (SNR) waveforms, scaled down to several magnitude levels and embedded in clean noise to derive models of measurement errors. We model these measurement errors as a series of generalized extreme value distributions whose parameters (location, scale, and shape) depend on the measured SNR. This allows us to model the increasing picking error bias and the increasingly heavier tails of the residual distributions with decreasing SNR.</p> <p>We have incorporated the full covariance matrix estimate in a linearized location algorithm. We show that by taking into account the correlated error structure with a robust transportable station-station correlation model, we achieve 90% coverage (i.e., the 90% error ellipse covers the true location 90% of the time), for sparse, unbalanced networks. Furthermore, the coverage statistics do not deteriorate with an increasing number of stations. Future work will incorporate SNR-dependent, non-Gaussian arrival-time measurement errors.</p>					
15. SUBJECT TERMS Seismic event detection, Seismic location, Seismic characterization, Seismic propagation					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT  SAR	18. NUMBER OF PAGES  9	19a. NAME OF RESPONSIBLE PERSON Robert J. Raistrick
a. REPORT UNCLAS	b. ABSTRACT UNCLAS	c. THIS PAGE UNCLAS			19b. TELEPHONE NUMBER (include area code) 781-377-3726

## IMPROVED EVENT LOCATION UNCERTAINTY ESTIMATES

István Bondár<sup>1</sup>, Keith McLaughlin<sup>1</sup>, and Hans Israelsson<sup>1</sup>

<sup>1</sup>Science Applications International Corporation

Sponsored by Air Force Research Laboratory

Contract No. FA8718-05-C-0018

### **ABSTRACT**

While many recent studies aimed to reduce location bias by introducing improved travel-time corrections, less effort was devoted to the complete estimation of location uncertainty, despite the fact that formal error ellipses are often overly optimistic. Since most location algorithms assume that the observations are independent, correlated systematic errors that are due to similar ray paths inevitably result in underestimated location uncertainties. Furthermore, the tails of real seismic data distributions are heavier than Gaussian. The main objectives of this project are to develop, test, and validate methodologies to estimate location uncertainties in the presence of correlated, systematic, and non-Gaussian errors. Particular attention is paid to robust and transportable models for a travel-time covariance matrix.

To address correlated errors, we estimate the spatial correlation structure in arrival-time data using variogram models. We developed a methodology based on copula theory to derive robust, data-driven variogram models. For validation purposes, we use GT0-2 event clusters. These include the Nevada, Lop Nor, Semipalatinsk, and Novaya Zemlya test sites, as well as the Azgir Peaceful Nuclear Explosions and the Lubin, Poland, mine-related events. Using ground-truth (GT) clusters allows us to calculate “ground truth” residuals with respect to the GT locations for a specific velocity model. We show the improvements in variogram estimates when using global 3D models instead of the iasp91 model. The proper choice of the underlying velocity model is especially important for regional phases.

To address issues embodied by real data distributions, we performed fully controlled experiments using known high signal-to-noise ratio (SNR) waveforms, scaled down to several magnitude levels and embedded in clean noise to derive models of measurement errors. We model these measurement errors as a series of generalized extreme value distributions whose parameters (location, scale, and shape) depend on the measured SNR. This allows us to model the increasing picking error bias and the increasingly heavier tails of the residual distributions with decreasing SNR.

We have incorporated the full covariance matrix estimate in a linearized location algorithm. We show that by taking into account the correlated error structure with a robust transportable station-station correlation model, we achieve 90% coverage (i.e., the 90% error ellipse covers the true location 90% of the time), for sparse, unbalanced networks. Furthermore, the coverage statistics do not deteriorate with an increasing number of stations. Future work will incorporate SNR-dependent, non-Gaussian arrival-time measurement errors.

DTIC COPY



## OBJECTIVES

The objectives of this project are to develop methodologies to estimate location uncertainties in the presence of correlated, systematic model errors and to characterize measurement errors as a function of signal parameters such as phase and SNR. The improved understanding of the complete error budget described by the full covariance matrix is incorporated into a linearized location algorithm, leading to more robust estimates of location uncertainty. The ultimate goal of this project is to develop transportable error models that will provide reliable location uncertainty estimates for small events recorded only by a few stations.

## RESEARCH ACCOMPLISHED

The motivation for this project is the observation that location uncertainty estimates are often underestimated, that is, the error ellipses scaled to the 90% confidence level do not contain 90% of the true locations. While most recent location calibration studies focused on producing improved travel-time predictions to reduce location bias, less effort was devoted to obtaining reliable formal error ellipses. Linearized location algorithms typically rely on the assumption that the observations are independent and the residuals are Gaussian distributed. Unfortunately, these two assumptions are almost always violated. Many researchers have pointed out (e.g., Buland, 1986; Anderson, 1982) that the distribution of measurement errors is non-Gaussian and better described by non-zero mean, skewed, heavy-tailed distributions. Furthermore, station distributions are far from uniform, and observations are not independent. Closely spaced stations sample similar ray paths and thus introduce correlated systematic errors. This redundancy in the observations reduces the effective number of degrees of freedom, thus ignoring the correlated structure inevitably results in underestimated location uncertainty estimates.

In this project we focus on the treatment of correlated errors combined with non-Gaussian, non-zero-mean, heavy-tailed, skewed distributions of reading errors. Figure 1 illustrates our research strategy. Both the measurement and model errors are described by their covariance matrices, where the full data covariance matrix ( $C_D$ ) is represented by the sum of the reading error ( $C_R$ ) and network ( $C_N$ ) covariance matrices. The network covariance matrix describes the correlated error structure and is estimated by variogram analyses. The reading errors are modeled by a general extreme value distribution, with parameters slowly increasing with decreasing SNR. The full data covariance matrix is incorporated in a linearized inversion scheme. Because of the non-Gaussian error distributions, the linearized inversion provides only an approximation of the solution; we will develop a hypothesis test to justify the validity of the "linearized" estimates.

Measurement error	Model error	Measurement error	Model error	Measurement error	Model error	
Gaussian	Uncorrelated	Gaussian	Correlated	Non-Gaussian	Correlated	
IID observations	IID observations	IID observations	Station-station variogram models	SNR-dependent bias and variance	Station-station variogram models	
Linearized inversion (diagonal covariance matrix) <b>BASELINE</b>		Linearized inversion with correlated errors (full covariance matrix)		Approximated by linearized inversion with correlated errors (full covariance matrix)		Copula-based hypothesis test for the validity of linear approach

Figure 1. Progress and strategy of research. Gray indicates the current state of the art linearized location algorithm. Validation test results are presented in this paper for the green stage (recent progress); yellow and orange represent planned future development.

All the methodologies developed in the course of this project are tested and validated using GT event clusters, including GT0-2 underground nuclear explosions from Yucca Flat and Pahute Mesa, Balapan and Degelen mountains, Novaya Zemlya, Azgir, and Lop Nor, as well as mining explosions from the Lubin mines.



### Network Covariance Matrix

The network covariance matrix accounts for the spatially correlated travel-time structure that is due to similar ray paths. We use variogram analysis to estimate the correlation structure in the data. In order to obtain robust variogram estimates, we use an entire GT event cluster. The variogram in geostatistical analysis is defined by the equation

$$\gamma(h) = \left\langle \left( \delta t(\Delta) - \delta t(\Delta + h) \right)^2 \right\rangle = \sigma_{\text{sill}}^2 - \text{Corr}(h) \sigma_{\text{sill}}^2 = \sigma_{\text{sill}}^2 - \text{Cov}(h), \quad (1)$$

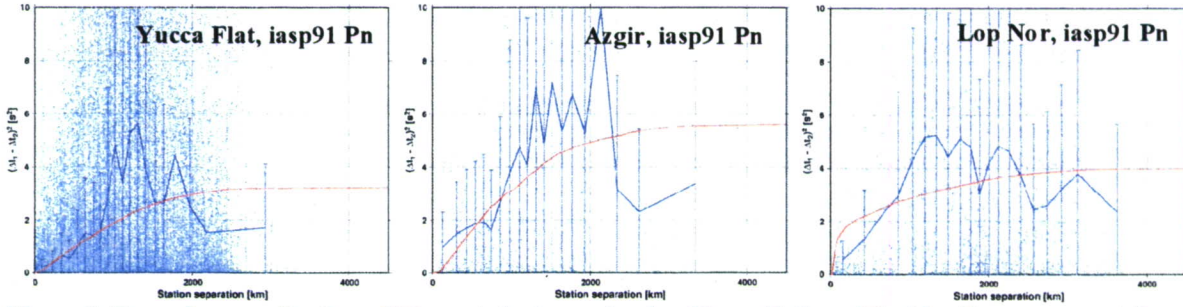
where  $\sigma_{\text{sill}}^2$  denotes the background variance,  $\delta t(\Delta)$  and  $\delta t(\Delta + h)$  are the GT residuals (residuals with respect to the GT location, using a specific Earth model) at station pairs of common events, and  $h$  is the station separation. We employ copula formalism (Nelsen, 1999; Frees and Valdez, 1998; Genest and Rivest, 1993) to determine robust variogram models. A copula couples marginal distributions to form the joint probability distribution:  $H(x, y) = C(F(x), G(y))$ . In other words, the copula describes the joint distribution of the order statistics  $u = F(h)$  and  $v = G\left(\left(\delta t(\Delta) - \delta t(\Delta + h)\right)^2\right)$ . We have evaluated a number of copulas and found that the Clayton copula, defined as  $C(u, v) = \left(u^{-\alpha} + v^{-\alpha} - 1\right)^{-1/\alpha}$ ,  $\alpha \in (0, \infty)$ , provides the best fit to the data (Bondár et al, 2005). We define the robust variogram model as the median regression curve of the residual difference squares for station pairs of common events with respect to station separation. Using the copula formalism, the median regression curve simply becomes the solution of the equation

$$C(v | u) = \frac{\partial C(u, v)}{\partial v} = 0.5, \text{ which yields } \gamma(h) = G^{-1} \left( \left[ 1 + u^{-\alpha} \left( 0.5^{-\alpha/(\alpha+1)} - 1 \right) \right]^{-1/\alpha} \right), u \in [0, 1]. \quad (2)$$

Thus, the copula formalism offers a fully data-driven approach to derive robust, monotonically increasing variogram models that are free from Gaussian assumptions and provide analytic expressions parameterized by the order statistics. Having derived a variogram model, the estimates for the elements of the network covariance matrix are

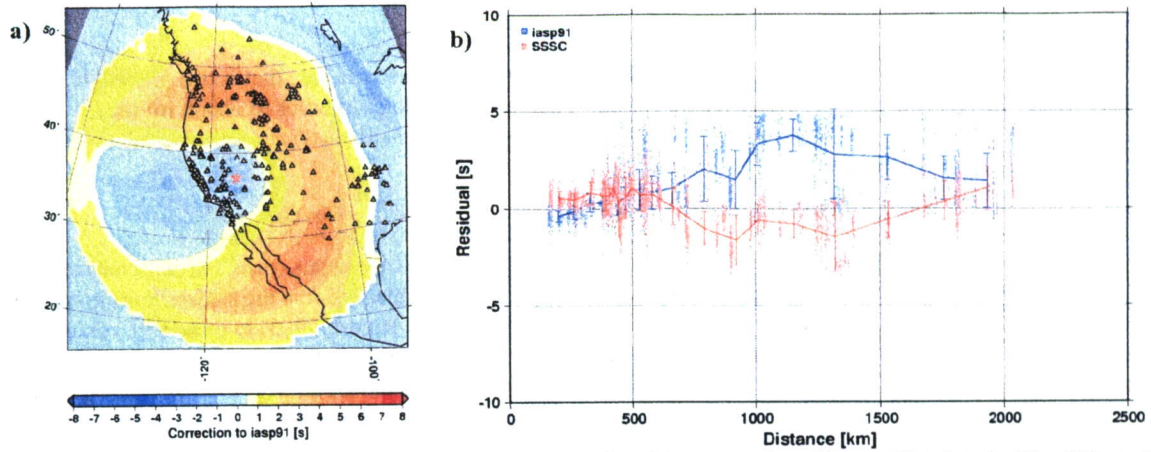
$$C_N(i, j) = \sigma_{\text{sill}}^2 - \gamma(\Delta(\text{sta}_i, \text{sta}_j)). \quad (3)$$

Note that the isotropic variogram model defined by Eq. (1) does not account for azimuthal variations in the correlation structure. Hence, the proper choice of the underlying velocity model to calculate residuals becomes important, especially for regional phases. Figure 2 shows the Pn variograms obtained for Yucca Flat, Azgir, and Lop Nor GT event clusters when using iasp91 (Kennett and Engdahl, 1991) predictions to calculate GT residuals. The variance of the residuals, indicated by the binned median (blue line), increases significantly at far regional distances, where stations from different tectonic provinces are mixed together. This is poorly modeled by the isotropic variogram model, shown by the red line.



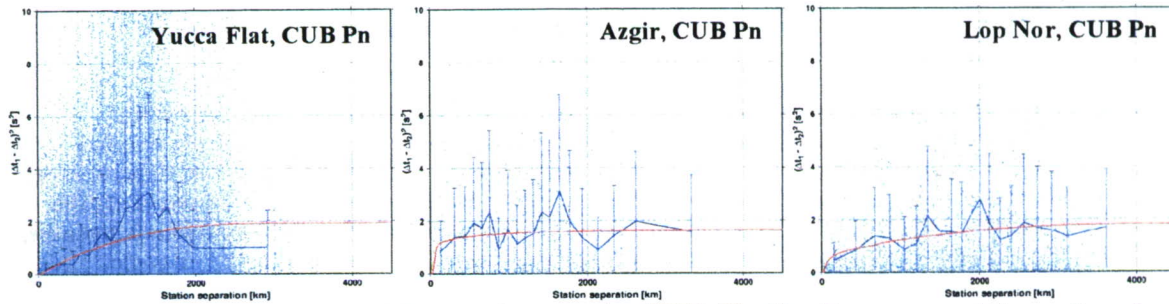
**Figure 2. Pn variograms for three GT event clusters using iasp91 predictions. The blue lines represent the median of every 5 percentiles; the red lines indicate the copula variogram models. Because of the unmodeled velocity structure by the iasp91 model, the isotropic variogram model provides a poor approximation of the correlation structure.**

To remedy these shortcomings, we recalculate the GT residuals using the global upper-mantle CUB2 model (Shapiro and Ritzwoller, 2004). The CUB2 model was extensively validated in Eurasia (Ritzwoller et al., 2003; Yang et al., 2004), but here we show an example for Pn observations from Yucca Flat events. Figure 3 shows the Pn travel-time correction surface relative to iasp91 predictions and the comparison of the iasp91 and CUB2 GT residuals as a function of epicentral distance. The CUB2 residuals are closer to zero mean than the iasp91 ones, indicating that the CUB2 significantly reduces the path effects that were left unmodeled by iasp91.



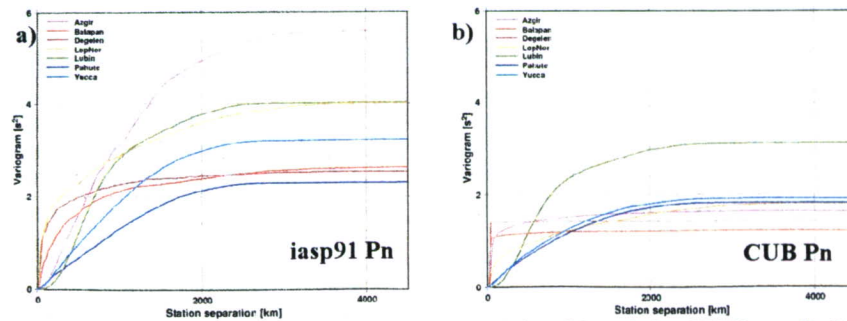
**Figure 3. (a)** CUB2 Pn travel-time correction surface to iasp91 centered on Yucca Flat (star). The 289 stations with Pn readings are shown as triangles. **(b)** GT residuals using iasp91 (blue) and CUB2 (red) as a function of epicentral distance.

Figure 4 shows the Pn variograms for the same event clusters as above, but now using the CUB2 predictions. The isotropic variogram models not only fit the observations better but also reflect a 50%–60% variance reduction typically achieved by the CUB2 model over iasp91.



**Figure 4. Pn variograms for three GT event clusters using CUB2.** The blue lines represent the median of every 5 percentiles, the red lines indicate the copula variogram models. Using a calibrated Earth model, the isotropic variogram models provide acceptable descriptions of the correlation structures.

Figure 5 shows the comparison of the Pn variogram models derived for the various GT event clusters when using iasp91 or CUB2 travel-time predictions. The variance reduction provided by the calibrated CUB2 predictions is reflected in the reduced and remarkably consistent background variance (sill).



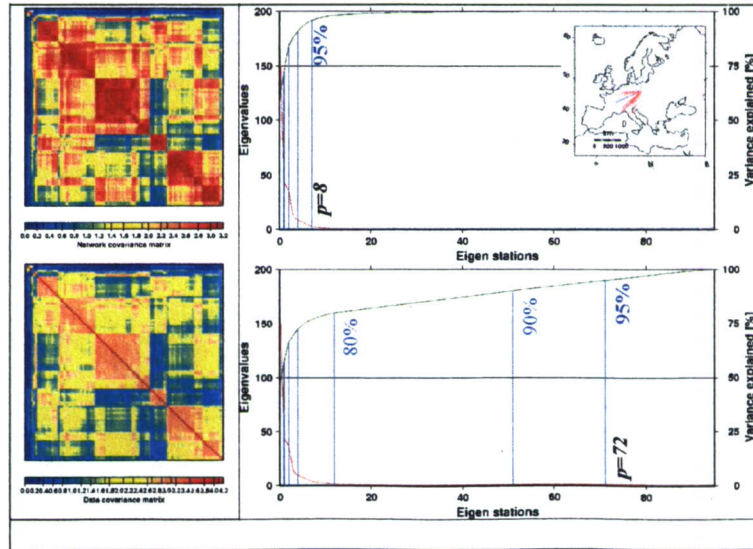
**Figure 5. Pn variogram models for GT event clusters using (a) iasp91 and (b) CUB2 predictions.** The CUB2 model produces much more consistent variograms, with reduced background variance (sill).



### Modified Location Algorithm

Standard linearized location algorithms assume independent Gaussian errors and solve the inversion problem by an iterative, weighted least-squares algorithm by minimizing the expression  $(d - Gm)^T C_R^{-1} (d - Gm)$ , which is equivalent to solving the equation  $WGm = Wd$ , where  $G$  is the  $(NxM)$  design matrix containing the travel time derivatives for an event-station path,  $m$  is the  $(Mx1)$  model adjustment vector  $[\Delta T, \Delta x, \Delta y, \Delta z]^T$ ,  $d$  is the  $(Nx1)$  vector of time residuals, and  $W = C_R^{-1/2}$  is the diagonal  $(NxN)$  weight matrix. The  $G_W m = d_W$  equation is often solved by singular value decomposition, which yields  $G_W^{-1} = V_W \Lambda_W^{-1} U_W^T$  and thus the model adjustment of  $m_{est} = G_W^{-1} d_W$ . At each iteration the model vector is adjusted such that  $m_{j+1} = m_j + m_{est}$ . Once a convergent solution is obtained, the location uncertainty is defined by the *a posteriori* model covariance matrix,  $C_M = G^{-1} C_R G^{-1T} = V_W \Lambda_W^{-2} V_W^T$ , which is typically scaled to the 90% confidence level. This linearized location algorithm constitutes our baseline, against which we measure improvements in location uncertainty estimates.

In the presence of correlated systematic errors, the data covariance matrix is no longer diagonal. Our modified linearized location algorithm seeks a transformed set of equations  $WGm = Wd$ , in which the data covariance matrix is diagonal. To ensure this, we solve the inversion problem in the eigen coordinate system in which the transformed observations are independent. The singular value decomposition of the full data covariance matrix is written as  $C_D = U_D \Lambda_D V_D^T$ , where  $\Lambda_D$  is the diagonal matrix of eigenvalues and the columns of  $U_D$  contain the eigenvectors of  $C_D$ . We keep the first  $p$  largest eigenvalues from the cumulative eigenvalue spectrum such that 95% of the total variance is explained:  $\sum_{j=1}^p \lambda_j / \sum_{i=1}^N \lambda_i \geq 0.95$ . We then define  $p$  as the effective number of degrees of freedom of the data, with an  $N-p$  dimension null space. Note that the 95% total variance level is somewhat arbitrary but a conservative and workable choice. Let  $C_D = BB^T$ , with  $B = U_p \Lambda_p^{1/2}$ , then the projection matrix  $W = B^{-1} = \Lambda_p^{-1/2} U_p^T$  orthogonalizes the data set and projects redundant observations into the null space. After applying the projections  $G_W = \Lambda_p^{-1/2} U_p^T G$  and  $d_W = \Lambda_p^{-1/2} U_p^T d$ , the formalism remains the same as in the baseline algorithm, but now  $d_W$  represents linear combinations of the observed residuals, the “eigen residuals.”



**Figure 6.** Upper panel: Network covariance matrix for a Lubin rockburst and its corresponding eigenvalue spectrum (red line). The green line shows the cumulative sum of eigenvalues as a percentage of the total sum. Only 8 eigenvalues are needed to explain 95% of the total variance. Bottom panel: Full data covariance matrix and its eigenvalue spectrum. Random picking errors blur the correlation structure and reduce the data redundancy.

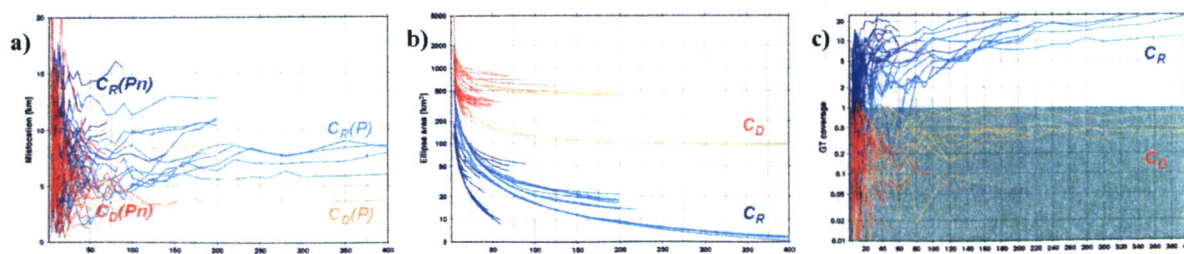
The process is illustrated for a Lubin, Poland, event that occurred on May 26, 1995. Pn was reported at 96 stations, representing a dense but rather unbalanced network, as shown in the Figure 6 inset. The upper panel of Figure 6 shows the network covariance matrix and its corresponding cumulative eigenvalue spectrum. The network covariance matrix has been arranged by its nearest-neighbor ordering of stations, and it exhibits a quasi-block-diagonal structure. Because many observations are strongly correlated, the first 8 largest eigenvalues explain 95% of the total covariance. In other words, 92% of the data carry redundant information. When the measurement (picking) error covariance matrix (assuming uniform, 0.5 s reading errors) is added to the network covariance matrix (lower panel), the reading errors weaken the correlation strength but leave the correlation pattern unchanged. Because of this “blurred” correlation structure, more eigenvalues are needed to explain 95% of the cumulative variance.

In general, when the picks suffer from large measurement errors, the correlation structure is blurred by the random noise of picking errors. However, when the onsets are picked very accurately, the correlation structure becomes very important for obtaining reliable location uncertainty estimates. Figure 7 shows location results with the assumption of independent observations compared with those accounting for correlated errors. When the correlation structure is accounted for, not only does the error ellipse cover the GT location but the mislocation is also reduced from 15.6 km to 5.5 km.

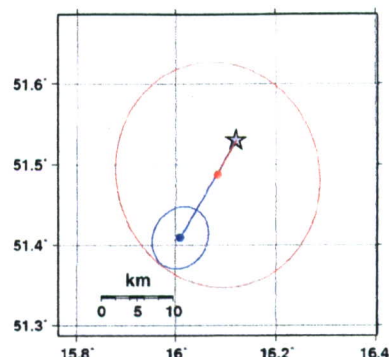
### Validation Tests

To validate the improved location uncertainty estimates that are due to the full data covariance matrix, we carried out two experiments using a set of events from each GT event cluster. In each experiment we located the events by both the baseline (independent error assumption) and the modified (correlated errors) location algorithms. We assume uniform, 0.5 s standard deviation measurement errors. For Pn we use the CUB2 travel-time model and construct the network covariance matrices from the variogram models obtained with CUB2 predictions. For teleseismic P we use iasp91 travel-time predictions.

The first experiment is designed to measure the robustness of the location uncertainty estimates against increasing levels of data redundancy. We locate events with an increasing number of stations of optimal azimuthal coverage and measure the mislocation, the area of the 90% error ellipse, and the GT coverage (whether the error ellipse contains the GT location). Figure 8 summarizes these results. When the correlated errors are accounted for, the improved relative weighting scheme reduces location errors. Since the inversion is performed in the eigensystem, redundant data are projected out and can no longer conspire to increase mislocation. The remaining mislocation represents the bias introduced by imperfect travel-time predictions. Furthermore, as the information content carried by the network is exhausted, the size of the 90% error ellipse stabilizes, while maintaining GT coverage. Thus, incorporating the correlated data structure in the location algorithm provides robust location uncertainty estimates unaffected by an increasing degree of redundant information.



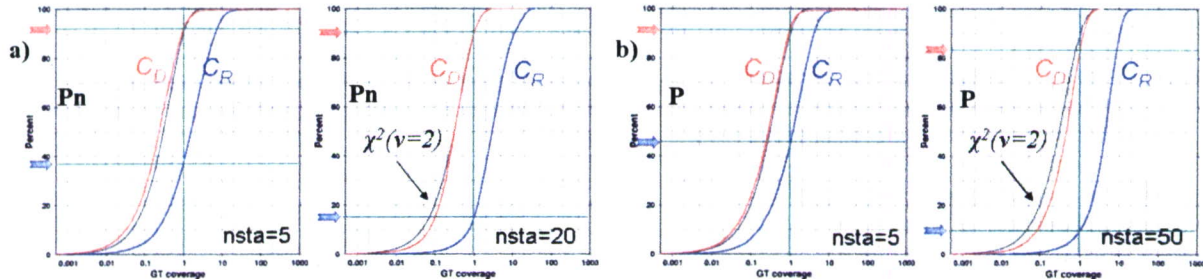
**Figure 8. (a) Mislocation, (b) error ellipse size, and (c) GT coverage with an increasing number of stations for an optimal network. Baseline location algorithm (independent error assumption) results are shown in blue; modified location algorithm (correlated errors) results are shown in red.**



**Figure 7. Relocation and 90% uncertainties of a Lubin GT1 event (star) using independence assumption (blue) and accounting for correlated errors (red).**



The second experiment is designed to measure the trustworthiness of the location uncertainty estimates for suboptimal networks. We locate events with randomly selected networks of 5, 8, 10, 20, and 50 stations (1000 realizations each) and measure the distribution of the coverage parameter (normalized distance inside vs outside the 90% ellipse). Note that under Gaussian assumptions, the coverage parameter follows a  $\chi^2$  distribution with 2 degrees of freedom. The actual GT coverage is the percentile value, where the coverage parameter equals to one. For honest 90% confidence error ellipses, this should obviously occur at the 90th percentile. Figure 9 shows the results for sparse and dense networks. While the assumption of independent observations produces abysmal coverage, taking into account the correlated error structure maintains 90% coverage for sparse, unbalanced networks. For dense networks, the coverage falls somewhat below 90%, and the deviation from the theoretical  $\chi^2$  distribution may indicate residual unmodeled nonlinear dependence structures or non-Gaussian effects.



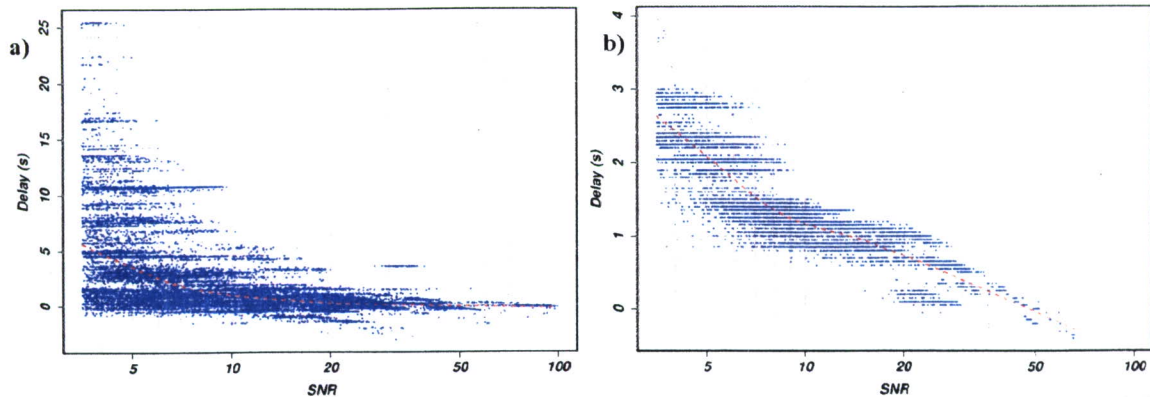
**Figure 9. (a) Pn and (b) P coverage statistics for suboptimal sparse and dense networks. The cumulative distribution of coverage parameters is shown in blue when events are located assuming independent errors and in red when using the full data covariance matrix. The theoretical distribution is drawn in black. When correlated errors are accounted for (red), the 90% confidence ellipse covers the true location for 90% of the realizations.**

### Preliminary Model of Measurement Errors

Measurement errors are typically modeled as Gaussian, zero-mean processes. However, picking errors of onset times suffer from heavy tails (Buland, 1986) and are skewed, as weak emergent arrivals are typically picked too late (Anderson, 1982). Measurement errors also suffer from systematic errors, as onset times along the same ray paths are systematically picked late with decreasing event size, or more precisely, with decreasing SNR (e.g., Douglas et al., 1997, 2005). Douglas et al. (2005) point out that automatic detections are more likely affected by the systematic errors than are manual picks made by experienced analysts. Systematic reading errors introduce location bias, especially for small events recorded by sparse, unbalanced regional networks.

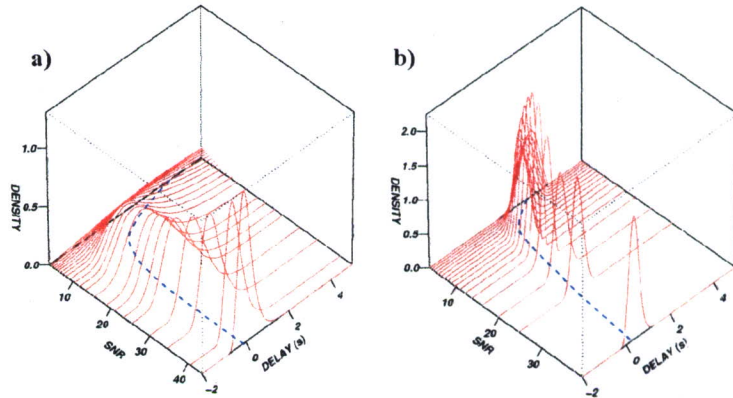
To develop models of measurement error bias and variance, we follow the methodology developed by Kohl et al (2004, 2005). The methodology scales known, high-SNR signals (explosions and/or earthquakes) down to various magnitude levels and embeds them in clean background noise, thus enabling large-scale controlled experiments. We generated some 22,000 realizations of earthquake waveforms from the Lop Nor region scaled to the range of mb 2.5–6.0, as well as some 9,000 realizations of Lop Nor explosion waveforms (mb 2.5–5.7), each embedded in different realizations of clean noise. Since the onset time and the SNR of the embedded signals are exactly known, running a signal detector allows us to derive improved measurement error models conditional on SNR. Figure 10 shows the delays of the automatic picks with respect to the true onset times for first arriving phases with SNRs larger than 3.5. It is apparent that the onsets are picked increasingly late, with decreasing SNR. Furthermore, the bias and the scatter around the bias are much smaller for the more-impulsive explosion signals, and picking errors on the earthquake signals exhibit heavier tails.





**Figure 10.** Delay of automatic first-arriving P picks relative to true onset times for (a) earthquake and (b) explosion waveforms. Picking errors of earthquake seismograms exhibit larger bias and scatter and suffer from heavier tails than those from explosions.

To model this increasing bias, scatter and skewness with decreasing SNR, we fit a series of general extreme value (GEV) distributions to the residual distribution as a function of SNR. Figure 11 illustrates the preliminary, SNR-dependent measurement error models for the earthquake and explosion picks. The location, scale, and shape parameters of the GEV distributions vary slowly with SNR, allowing for continuous change in reading errors. The increasing reading error bias with decreasing SNR (dashed lines in Figure 11) introduces systematic, correlated picking errors for a fixed event. However, once the bias is corrected for, the reading errors become nearly independent; thus, the reading error covariance matrix may still be approximated by a diagonal matrix.



**Figure 11.** Measurement error models for (a) earthquake and (b) explosion signals for first-arriving P presented as a series of GEV distributions, with parameters slowly changing with SNR.

## CONCLUSIONS AND RECOMMENDATIONS

We have developed methodologies to obtain robust estimates of network covariance matrices through copula-based variogram models. We have shown that the CUB2 global upper-mantle model provides significant variance reduction over the iasp91 model for regional Pn and thus produces reliable isotropic variogram models. We have developed a modified linearized location algorithm to incorporate the full, non-diagonal data covariance matrix in the inversion problem. Validation tests demonstrated that when the correlated data structure is taken into account, we

- obtain robust location uncertainty estimates, unaffected by the amount of redundant information,
- reduce mislocation that is due to conspiring errors, and
- achieve 90% coverage 90% of the time for both sparse and dense unbalanced networks.

While the variogram models for the various GT event clusters produce slightly different covariance matrices, the overall correlation structures are similar as the correlation varies the most at relatively small station separations (up to a few hundred kilometers). Preliminary tests indicate that it is possible to develop generic transportable variogram models for both Pn and P that will err on the conservative side and may work well anywhere on the globe.

We have developed preliminary measurement error models for first-arriving phases. We model the bias and scatter of picking errors by a series of GEV distributions whose parameters change slowly with decreasing SNR. In the next stage of this project, we will incorporate these non-Gaussian picking error models into our modified linearized location algorithm, hence utilizing a complete covariance matrix based on spatially correlated and SNR dependent errors. A hypothesis test will then be developed to verify when the linearized approach is valid.

## REFERENCES

- Anderson, K. R. (1982). Robust earthquake location using M estimates, *Phys. Earth Planet. Int.* 30: 119–130.
- Bondár, I., K. McLaughlin, and H. Israelsson (2005). Improved event location uncertainty estimates, in *Proceedings of the 27th Seismic Research Review: Trends in Nuclear Explosion Monitoring*, LA-UR-05-6407, Vol 1, pp. 299–307.
- Buland, R. (1986). Uniform reduction error analysis, *Bull. Seism. Soc. Am.* 76: 217–230.
- Douglas, A., D. Bowers, and J.B. Young (1997). On the onset of P seismograms, *Geophys. J. Int.* 129: 681–690.
- Douglas, A., J. B. Young, D. Bowers, and M. Lewis (2005). Variation in reading error in P times for explosions with body-wave magnitude, *Phys. Earth Planet. Int.* 152: 1–6.
- Frees, E. W. and E. A. Valdez (1998). Understanding relationships using copulas, *North Am. Actuarial J.* 2: 1–25.
- Genest, C. and L-P. Rivest (1993). Statistical inference procedures for bivariate Archimedean copulas, *J. Am. Stat. Assoc.* 88: 1034–1043.
- Kennett, B. L. N. and E. R. Engdahl (1991). Travel times for global earthquake location and phase identification, *Geophys. J. Int.* 105: 429–465.
- Kohl, B., T. J. Bennett, I. Bondár, B. Barker, W. Nagy, C. Reasoner, and J. Hanson (2004). Development of a network data set for evaluating detection and network processing performance, in *Proceedings of the 26th Seismic Research Review: Trends in Nuclear Explosion Monitoring*, LA-UR-04-5801, Vol. 2, pp. 725–734.
- Kohl, B., T. J. Bennett, I. Bondár, B. Barker, W. Nagy, C. Reasoner, H. Israelsson, and P. Piraino (2005). Development of a network data set for evaluating detection and network processing performance, in *Proceedings of the 27th Seismic Research Review: Trends in Nuclear Explosion Monitoring*, LA-UR-05-6407, Vol. 2, pp. 917–926.
- Nelsen, R. B. (1999). *An Introduction to Copulas*. Lecture Notes in Statistics 139, New York: Springer Verla.
- Ritzwoller, M. H., N. M. Shapiro, E. A. Levshin, E. A. Bergman, and E. R. Engdahl (2003). Ability of a global three-dimensional model to locate regional events, *J. Geophys. Res.* 108(B7): 2353, 10.1029/2002JB002167.
- Shapiro, N. M. and M. H. Ritzwoller (2004). Thermodynamic constraints on seismic inversions, *Geophys. J. Int.* 157: 1175–1188, doi:10.1111/j.1365-246X.2004.02254.x.
- Yang, X., I. Bondár, J. Bhattacharyya, M. Ritzwoller, N. Shapiro, M. Antolik, G. Ekström, H. Israelsson, and K. McLaughlin (2004). Validation of regional and teleseismic travel-time models by relocating GT events, *Bull. Seism. Soc. Am.* 94: 897–919.